DOCUMENT RESUME

ED 326 576 TM 015 938

AUTHOR Olsen, Scott A.
TITLE Examining the Relationship between the CAAP Critical Thinking Test and the COMP.
PUB DATE Apr 90
NOTE 52p.; Paper presented at the Annual Meeting of the American Educational Research Association (Boston, MA, April 16-20, 1990).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Achievement Gains; *Achievement Tests; *Cognitive Tests; College Freshmen; *College Sophomores; Comparative Testing; *Critical Thinking; Higher Education; Outcomes of Education; Scores; Test Results
IDENTIFIERS American College Testing Program; *College Outcome Measures Project; *Collegiate Assess Acad Prof Crit Think Test; Testing Effects; Test Order

ABSTRACT

This study examined the relationship between the Collegiate Assessment of Academic Proficiency Critical Thinking test (CAAP-CT) and the College Outcome Measures Program objective test battery (COMP). Focus was on determining: (1) the relationship between the CAAP-CT and the COMP process-oriented subscores and the freshman to sophomore gains in these subscores; and (2) at a composite score level, the relationship of the CAAP-CT score with the COMP total, COMP total gain, entering American College Testing (ACT) Program composite score, and cumulative grade point average. A secondary objective was to determine whether or not a test order effect exists within an approximately 3.5-hour testing session, by comparing those who took the CAAP-CT followed by the COMP with those who took the COMP followed by the CAAP-CT. Data were obtained at Northeast Missouri State University (Kirksville) from students taking their required sophomore examinations during the fall of 1988 and the spring of 1989. The data set analyzed contained valid observations for 568 students. Results indicate that the order in which students took the COMP and CAAP-CT had no significant impact on the score results. There are significant zero-order correlations between COMP process-oriented subscores and CAAP-CT scores. The COMP-based process subscores account for at least 33% of the variance in CT and up to 40% with COMP-based gain scores included. There was insufficient evidence to support the contention that higher gains are associated with higher levels of CT. The ACT maintained a greater relationship to the CAAP-CT than did the COMP, which was administered at the same time and offers more tasks requiring CT skills. Twenty-three data tables are included. (RLC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED326576

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Scott A. Olsen

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Examining the Relationship Between the CAAP Critical Thinking Test and the COMP

Scott A. Olsen
Northeast Missouri State University

Paper presented in Division J, session 18.17, of the American Educational Research Association's annual conference in Boston, MA, April 16-20, 1990.

TM015938

INTRODUCTION

The development and assessment of critical thinking skills has long been a prominent topic in discussions of higher education and, in recent years, it has received increased emphasis. Moreover, in a recent review of critical thinking studies, James McMillan (1987) cites evidence that this emphasis on critical thinking skills is likely to increase. For example, he points to Integrity in the College Curriculum: A Report to the Academic Community (Association of American Colleges, 1985) which proposes that students learn to utilize inquiry, logic, critical analysis and other reasoning skills to subject ideas, data and other forms of evidence to the "probing analysis of the mind." And he quotes Involvement in Learning: Realizing the Potential of American Higher Education (National Institute of Education, 1984) as recommending that curriculum promote "the development of capacities of analysis, problem solving, communication, and synthesis" and "the ability to think critically, to synthesize large quantities of new information," as well as, that "faculty and academic deans ... design and implement a systematic program to assess the knowledge, capacities, and skills developed by students ... [and] ... it is especially important that these assessments allow for judgments of the impact ... on analytic reasoning, and the ability to synthesize."

Many definitions of critical thinking and what it involves exist. Robert Ennis, a prominent American specialist in the area of critical thinking, defines critical thinking as "reasonable reflective thinking that is focused on deciding what to believe or do" and provides thirteen dispositions and twelve abilities to aid in the identification of critical thinking (Ennis, 1985). Glaser (1985) indicates that the ability to recognize assumptions and

values, evaluate arguments and evidence, and draw sound inferences are key components to critical thinking. In addition to recognizing assumptions, Brookfield (1987) contends that critical thinking involves challenging assumptions and context to explore alternatives in a way that can lead to reflective skepticism. Moore and Parker (1989) in their second edition of Critical Thinking: Evaluating Claims and Arguments in Everyday Life, define critical thinking as "the careful and deliberate determination of whether to accept, reject, or suspend judgment about a claim." Postman (1985) points out that critical thinking involves more than the analysis of word-bound information in the electronic age of auditory and visual images. Sternberg (1985) suggests that problem-solving ability may be one way to conceptualize critical thinking.

A recently constructed test of critical thinking specifically designed for college level students is the Collegiate Assessment of Academic Proficiency (CAAP) Critical Thinking test (CT) developed by American College Testing (ACT) programs. The Report on the Technical Characteristics of CAAP: Pilot Year 1 1988-89, describes CT as a 40 minute test comprised of 32 items measuring "the ability to clarify, analyze, evaluate and extend arguments" where argument is defined as "a sequence of statements which includes a claim that one of the statements, the conclusion, follows from the other statements" (ACT, 1989). This report further states that

> "The test consists of four passages that are representative of the kinds of issues commonly encountered in a postsecondary curriculum. A passage typically presents a series of subarguments in support of a more general conclusion or conclusions. Each passage presents one or more arguments and uses a variety of formats, including case studies, debates, dialogues, overlapping positions, statistical arguments, experimental results, or editorials. Each passage is accompanied by a set of multiple-choice test items." (p. 13)

The report goes on to describe the content of the CT as analysis of the elements of arguments (62%), evaluation of arguments (19%) and extension of

arguments (19%). The CT is reported to have a reliability of between .81-.82 (KR-20), a standard error of measurement of 2.3 for scaled scores that have a maximum range of 45 to 75.

The College Outcome Measures Program (COMP) objective test battery, also developed by ACT, has been widely used for a number of years by colleges and universities across the country for program assessment. Table 1 gives a description of the COMP outcome areas measured, which are divided into process and content outcomes. A subscore is provided for each of the six outcome areas. The COMP "Technical Report 1976-1981" (Forrest & Steele, 1982) reports Cronback alpha reliability estimates of 0.84 for the COMP Total score, 0.63 (FSI), 0.68 (US), and 0.66 (UA) respectively for the content subscores, and 0.66 (COM), 0.68 (SP), and 0.66 (CV) respectively for the process subscores. Including directions and closure time, the COMP objective test battery takes 150 minutes to administer.

Comparing the descriptions given in Table 1 with the components of critical thinking definitions given earlier, it would appear that the process areas of the COMP may address some of the dimensions of what is called critical thinking. Furthermore the ability to clarify, analyze, evaluate and extend arguments measured by the CAAP CT would seem to involve the types of thinking necessary do well in the COMP process areas. Indeed a review of the items on both tests suggests a fair degree of overlap in the type of thinking required to respond to a number of items. Thus, while not designed as a critical thinking test, the COMP would seem to address some of those skills referred to as critical thinking skills.

Table 1
COMP Outcome Areas

Process Areas

Communication (COM): Can send and receive information in a variety of modes (written, graphic, oral, numeric, and symbolic), within a variety of settings (one-to-one and in small and large groups), and for a variety of purposes (for example, to inform, to understand, to persuade, and to analyze).

Solving Problems (SP): Can analyze a variety of problems (for example, scientific, social, and personal); select or create solutions to problems; and implement solutions.

Clarifying Values (CV): Can identify one's personal values and the personal values of other individuals; understand how personal values develop; and analyze the implications of decisions made on the basis of personally held values.

Content Areas

Functioning Within Social Institutions (FSI): Can identify those activities and institutions which constitute the social aspects of a culture (for example, governmental and economic systems, religion, marital and familial institutions, employment, and civic volunteer and recreational organizations); understand the impact that social institutions have on individuals in a culture; and analyze one's own and others' personal functioning within social institutions.

Using Science and Technology (US): Can identify those activities and products which constitute the scientific/technological aspects of a culture (for example, transportation, housing, energy, processed food, clothing, health maintenance, entertainment and recreation, mood alteration, national defense, communication, and data processing); understand the impact of such activities and products on the individuals and the physical environment in a culture; and analyze the uses of technological products in a culture including one's personal use of such products.

Using the Arts (UA): Can identify those activities and products which constitute the artistic aspects of a culture (for example, graphic art, music, drama, literature, dance, sculpture, film and architecture); understand the impact that art, in its various forms, has on individuals in a culture; and analyze uses of works of art within a culture and one's personal use of art.

* All outcome area descriptions were obtained from ACT-COMP materials (1987).

Many colleges and universities across the country use COMP gain scores as one measure of growth related to their academic programs. Such gain scores typically are formed by administering the COMP twice and subtracting the first score, typically obtained as students are entering as freshmen, from the second score, typically obtained at the end of the sophomore year (the gain covering the general education or core portion of the curriculum) or at the end of the senior year (the gain covering the entire program for a discipline or university). Since college level programs usually intend to promote critical thinking as one outcome, and since COMP gains from time one to time two are often employed as one measure of group growth, one could speculate that higher COMP gain scores might be related to higher CT scores, particularly for the process-oriented subscore gains.

Objectives

The primary objective of this study is to examine the relationship between the CAAP Critical Thinking test and the COMP objective test. This objective will be addressed at two levels. First, the relationship between CT and the COMP process-oriented subscores as well as the freshman to sophomore gains in these subscores will be investigated. Secondly, at a composite score level, the relationship of CT with COMP Total, COMP TOTAL-Gain, entering ACT composite score and cumulative grade point average, will be examined.

A secondary objective arises in this study both as a consequence of the design employed and as a prerequisite to the primary objective stated above. This secondary objective is to determine if there exists a test order effect within an approximately three and one-half hour testing session, by comparing those who took the CT followed by the COMP with those who took the COMP followed by the CT.

METHODOLOGY

Data Source

All data for this study were obtained at Northeast Missouri State University (NMSU) from students taking their required sophomore examinations during the fall of 1988 and the spring of 1989. NMSU is a selective medium sized public liberal arts and sciences state university. The university's total enrollment is around 7000 undergraduate and graduate students. From an applicant pool of over 7500, between 1600 and 1700 freshmen are admitted each year. The entering freshmen class American College Testing (ACT) assessment battery composite score mean ranges between 23 and 24.5.

The university has a long history of outcome assessments starting in 1974. Since 1983 NMSU has used the College Outcome Measures Program (COMP) test as one part of its assessment program. In this part of the assessment program, NMSU requires that all students take specified tests measuring general educational development when they come in as freshmen and again at the end of their sophomore year (defined as between 44 and 76 semester hours of credit, and students must take their sophomore examinations before they can enroll for hours beyond 76). Fifty percent of the freshmen are required to take the COMP each year.

Those students who were re-administered the COMP as sophomores during the fall of 1988 and spring of 1989, were also administered the Collegiate Assessment of Academic Proficiency (CAAP) Critical Thinking (CT) test. Half of those scheduled for sophomore examinations took the COMP test immediately before the CT test, while the other half took the CT test immediately before the COMP test, thereby counterbalancing test order. Within each semester the assignment to test order was done using a nine cell stratified random design controlling for three levels of entering ACT composite scores (approximate

high, middle and low thirds of the distribution) in combination with three levels of total credit hours earned (approximate high, middle and low thirds of the distribution). All testing sessions were approximately three and one-half hours in length. Thus the data used in this study are from required test administration (not voluntary) in a stratified random assignment design controlling for college entry level of ability and total college credit hours at time of testing, and with test order counterbalanced.

Variables

Table 2 provides definitions for the primary variables used in this study. While most of these variables require no additional explanation, elaboration on the formation of some of the COMP-based variables seems warranted.

First only the Process subscores and not the Content subscores of the COMP are used. There are two reasons for this. First as indicated in the introduction, a review of the COMP test content together with logic suggests that Process subscores would be most likely related to CT. The second reason is technical, in that the COMP scoring procedure makes the Content and Process sets of scores dependent upon one another. Hence, utilization of only the Process scores eliminates undesirable artifacts in the exploration of relationships, consistent with the purposes of this study. For similar reasons the COMP Total score, which is essentially a composite of the process subscores, will be examined in separate analyses from analyses of subscore with CT relationships.

Table 2
Variable Definitions

Entering Freshmen Measures

ACT = Entering Freshmen ACT assessment battery composite score

Freshmen COMP Scores:
- FTOTAL = COMP Total composite score from the objective test battery
- FSP = Solving Problems subscore
- FCV = Clarifying Values subscore
- FQCOM = Quantitative Communications portion of the communications subscore
- FVCOM = Verbal Communications part of the communications subscore

End of Sophomore Year Measures

CT = CAAP Critical Thinking test score

Sophomore COMP Scores:
- TOTAL = COMP Total composite score from the objective test battery
- SP = Solving Problems subscore
- CV = Clarifying Values subscore
- QCOM = Quantitative Communications portion of the communications subscore
- VCOM = Verbal Communications portion of the communications subscore

Freshmen to Sophomore COMP Gain Scores:
- TOTAL-GAIN = COMP TOTAL-Gain score defined as TOTAL-FTOTAL
- SP-GAIN = Solving Problems Gain score defined as SP-FSP
- CV-GAIN = Clarifying Values Gain score defined as CV-FCV
- QCOM-GAIN = Quantitative Communications Gain score defined as QCOM-FQCOM
- VCOM-GAIN = Verbal Communications Gain score defined as VCOM-FVCOM

CREDITS = Total semester hours of credit earned at the time of sophomore examinations

GPA = Cumulative grade point average on a 4-point scale at the end of the sophomore examination academic year.

AGE = Age at the time of sophomore examination.

ORDER = Order in which sophomore examinations were administered within the testing session (CT-COMP or COMP-CT)

SEMESTER = Semester of sophomore examination (Fall 1988 or Spring 1989)

Secondly, as seen in Table 2 the COMP communication subscore has been broken down into two parts, quantitative communication (QCOM) and verbal communication (VCOM), for lack of better titles. QCOM is an experimental score provided by ACT that assesses ability to communicate via numeric, graphic, and symbolic computational activity. The QCOM score is based upon the eight communication items that require computations. VCOM is the score based upon the rest of the items leading to the communication subscore normally reported. VCOM is derived simply by subtracting the QCOM subscore from the COMP communication subscore.

Finally gain scores were formed in the usual manner. Specifically, each COMP gain score was computed by subtracting the freshman score from the sophomore score. While such scores hold meaning in group-based program analyses, it should be recognized that such gain scores are not sufficiently reliable for individual student interpretations.

Sample Description and Data Integrity

The analysis of data integrity is important in studies of this kind. First students often fail to show for their scheduled testing session for various reasons including such things as conflicting commitments (university or personal), illness, oversight, suspension, dropping out of school, testing waiver (e.g. disabled), etc. Secondly, for those who do show for their scheduled testing session, experience indicates that a small yet potentially problematical proportion of students will produce invalid scores which can be detected and eliminated. In either case, comparison of those who were retained in the "sample" and those who were not is logically warranted in the description of the final data base to be used.

In this study student records were coded invalid and subsequently deleted for the following reasons:

1. The student did not show up for his/her scheduled sophomore examination and therefore had missing scores.
2. The student took the sophomore examination early and therefore fell outside of the 1988-89 academic year time frame for the study.
3. The student had a below chance score (130) on the freshmen COMP test.
4. The student had a below chance score (130) on the sophomore COMP test.
5. The student had an unrealistic freshman to sophomore COMP TOTAL-Gain of over 40 points.
6. The student had an unrealistic freshman to sophomore COMP Total loss (negative gain score) of 10 (-10) points or more.
7. The student had a COMP Total score 20 points or more below the ACT assessment battery composite score to COMP Total score concordance expectation.

Reasons 3-7 on this list are those suggested and encouraged by ACT for those using the COMP testing program (COMP, 1990). It should be noted that often a student's data record will be deemed invalid for more than one reason. Table 3 documents the number of student records deleted from the sample within each semester of sophomore examinations while specifying the number of reasons each record was found to be invalid. As Table 3 shows, some (N=7) students were deleted for as many as four reasons while 568 students were retained for analysis in this study.

One may also notice from Table 3, that 72.4% of the scheduled sophomore examinations were for the spring semester. This is not unusual since to qualify for sophomore examinations students must have between 44 and 74 semester hours of credit and are typically scheduled for examinations the first semester after they have entered this range. This window naturally tends to present a very skewed distribution of credit hours due to the fact that most students take around 15-16 hours of credit each semester and take

Table 3
Number of Valid and Invalid Student Records
Summarized by Semester and Number of Reasons for Invalidity

INVALID	SEMESTER		
FREQUENCY / PERCENT / ROW PCT / COL PCT	Fall	Spring	TOTAL
0 Reasons (Thus Retained)	159 22.05 27.99 79.90	409 56.73 72.01 78.35	568 78.78
1 Reason	29 4.02 28.71 14.57	72 9.99 71.29 13.79	101 14.01
2 Reasons	6 0.83 40.00 3.02	9 1.25 60.00 1.72	15 2.08
3 Reasons	5 0.69 16.67 2.51	25 3.47 83.33 4.79	30 4.16
4 Reasons	0 0.00 0.00 0.00	7 0.97 100.00 1.34	7 0.97
TOTAL	199 27.60	522 72.40	721 100.00

sophomore examinations at the end of their sophomore year in their fourth semester. Table 4 describes the distribution of total semester hours of credit at time of sophomore examination for the 568 students who had valid records and who were consequently retained in the sample.

Table 4
Distribution of Total Semester Hours of Credit

MOMENTS

N	568	SUM WGTS	568
MEAN	49.9329	SUM	28361.9
STD DEV	5.10092	VARIANCE	26.0194
SKEWNESS	1.02846	KURTOSIS	0.686921
USS	1430946	CSS	14753
CV	10.2156	STD MEAN	0.21403

RANGE	30
Q3-Q1	7
MODE	46

QUANTILES(DEF=4)

100%	MAX	74	99%
75%	Q3	53	95%
50%	MED	48	90%
25%	Q1	46	10%
0%	MIN	44	5%

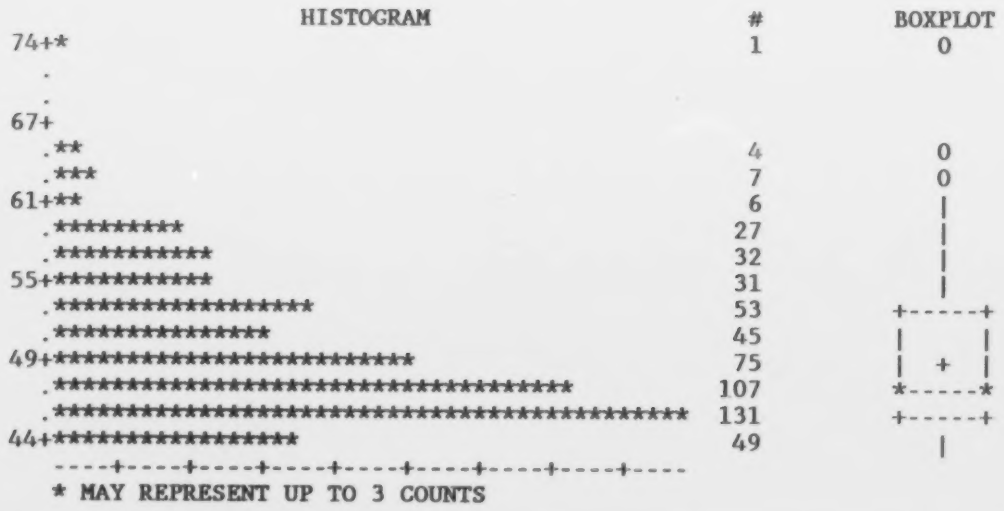

Table 5 summarizes the proportion of students missing within each of the assigned test order classifications. There were 721 students scheduled to take sophomore examinations with 37 (5.13%) who did not show. There was no significant difference (P=0.423) between proportions missing in the COMP-CT order as compared to the CT-COMP order.

Table 5
Frequency of Those Missing Their Scheduled Sophomore Examinations
Compared by Testing Order Assigned

MISSING	ORDER		
FREQUENCY PERCENT ROW PCT COL PCT	COMP-CT	CT-COMP	TOTAL
NO	342	342	684
	47.43	47.43	94.87
	50.00	50.00	
	94.21	95.53	
YES	21	16	37
	2.91	2.22	5.13
	56.76	43.24	
	5.79	4.47	
TOTAL	363	358	721
	50.35	49.65	100.00

CHI-SQUARE (DF=1) = 0.641, P=0.423

In Table 6 entering ACT composite score, freshmen COMP Total score, total semester hours of college credit earned at the time of sophomore examination, and cumulative grade point average (4-point scale) at the end of the academic year in which sophomore examinations were taken, are used to compare those who missed with those who did not miss their scheduled sophomore examinations.

Table 6
Those Missing Their Scheduled Sophomore Examinations
Compared to Those Not Missing

ENTERING ACT COMPOSITE SCORE

MISSING	N	MEAN	STD DEV	STD ERROR
NO	683	24.42459736	3.76993562	0.14425258
YES	37	23.94594595	3.55860417	0.58503092

FOR HO: VARIANCES ARE EQUAL, F'(682,36) = 1.12 PROB > F' = 0.6920
FOR HO: MEANS ARE EQUAL T'(718) = 0.7543 PROB > |T'|= 0.4509

FRESHMAN COMP TOTAL SCORE

MISSING	N	MEAN	STD DEV	STD ERROR
NO	684	186.69590643	14.17257929	0.54190211
YES	37	186.86486486	14.38008299	2.36407108

FOR HO: VARIANCES ARE EQUAL, F'(36,683) = 1.03 PROB > F' = 0.8468
FOR HO: MEANS ARE EQUAL T'(719) = -0.0706 PROB > |T'| = 0.9438

TOTAL SEMESTER HOURS OF COLLEGE CREDIT

MISSING	N	MEAN	STD DEV	STD ERROR
NO	684	49.80394737	5.03745551	0.19261193
YES	37	52.23243243	9.88259002	1.62468779

FOR HO: VARIANCES ARE EQUAL, F'(36,683) = 3.85 PROB > F' = 0.0001
FOR HO: MEANS ARE EQUAL T'(37) = -1.4843 PROB > |T'|= 0.1462*
* Adjusted for unequal variances

CUMULATIVE GRADE POINT AVERAGE

MISSING	N	MEAN	STD DEV	STD ERROR
NO	684	2.95719298	0.53182359	0.02033478
YES	37	2.56864865	0.55444748	0.09115060

FOR HO: VARIANCES ARE EQUAL, F'(36,683) = 1.09 PROB > F' = 0.6739
FOR HO: MEANS ARE EQUAL T'(719) = 4.3191 PROB > |T'|= 0.0001

The null hypotheses that the variances are equal and that the means are equal were tested for each variable.

As Table 6 indicates, there were no significant differences in variances or means on the first two variables, ACT composite scores and freshmen COMP Total scores. Thus, with respect to entering freshmen ability levels, the 37 who missed sophomore examinations did not differ from the others.

With respect to the sophomore variables, total credit hours at the time of scheduled sophomore examinations and cumulative GPA, there were two differences that could have been anticipated. First, there was a difference in variances for total hours of credit. And secondly, there was a difference in cumulative GPA means.

The difference in variances was probably in part a function of the window of total credit hours used to define qualification for sophomore exams (44 to 74 hours) and the skewness of the resulting distribution--see Table 4. But in addition, many of the reasons for missing sophomore examinations (e.g. suspension, illness, dropping out or transferring) after qualifying for them would seem to be related to a wider range of earned credit hours, though not necessarily to a difference in means. For instance those suspended after three semesters may have just passed the 44 credits mark yet they would not be enrolled the semester of sophomore examination, whereas students transferring out of the university might put off taking sophomore examinations until after two years at NMSU (60 hours or more) and then would leave.

The difference in mean cumulative GPA is not surprising since many of those who do not show for their scheduled sophomore examination are dropping out of school or have been suspended for academic reasons. Such students are obviously most likely to have low GPA's, which will pull the average down.

Of those students who took their sophomore examinations as scheduled (N=684) there were 116 that were deleted for one or more of the validity reasons listed as 2-7 at the beginning of the section. Table 7 compares the number of valid COMP records for the two testing orders (P=0.839).

Table 7
Valid and Invalid Records for Non-missing Students
Compared by Testing Order

VALID	ORDER		
FREQUENCY / PERCENT / ROW PCT / COL PCT	COMP-CT	CT-COMP	TOTAL
NO	59	57	116
	8.63	8.33	16.96
	50.86	49.14	
	17.25	16.67	
YES	283	285	568
	41.37	41.67	83.04
	49.82	50.18	
	82.75	83.33	
TOTAL	342	342	684
	50.00	50.00	100.00

CHI-SQUARE (DF=1) = 0.042, P=0.839

Table 8 compares those with valid COMP records to those with invalid COMP records on relevant non-COMP-based variables, CT, CREDITS, GPA, and ACT. No significant differences in either variances or means were found between valid and invalid COMP records.

Table 8
Comparison of Valid and
Invalid Records for Non-missing Students

CRITICAL THINKING SCORE

VALID	N	MEAN	STD DEV	STD ERROR
NO	114	66.19298246	3.93279201	0.36833972
YES	541	66.75046211	3.69817582	0.15899700

FOR HO: VARIANCES ARE EQUAL F'(113,540) = 1.13 PROB > F' = 0.3770
FOR HO: MEANS ARE EQUAL T'(653) = -1.4465 PROB > |T'| = 0.1485

TOTAL HOURS OF CREDIT

VALID	N	MEAN	STD DEV	STD ERROR
NO	116	49.17241379	4.68397056	0.43489575
YES	568	49.93292254	5.10092400	0.21403002

FOR HO: VARIANCES ARE EQUAL F'(567,115) = 1.19 PROB > F' = 0.2610
FOR HO: MEANS ARE EQUAL T'(682) = -1.4830 PROB > |T'| = 0.1385

CUMULATIVE GPA

VALID	N	MEAN	STD DEV	STD ERROR
NO	116	2.97879310	0.50714059	0.04708682
YES	568	2.95278169	0.53704890	0.02253407

FOR HO: VARIANCES ARE EQUAL F'(567,115) = 1.12 PROB > F' = 0.4534
FOR HO: MEANS ARE EQUAL T'(682) = 0.4798 PROB > |T'| = 0.6315

ENTERING ACT COMPOSITE SCORE

VALID	N	MEAN	STD DEV	STD ERROR
NO	115	24.98260870	3.60307522	0.33598850
YES	568	24.31161972	3.79590499	0.15927264

FOR HO: VARIANCES ARE EQUAL F'(567,114) = 1.11 PROB > F' = 0.4982
FOR HO: MEANS ARE EQUAL T'(681) = 1.7432 PROB > |T'| = 0.0818

From the N's given in Table 8, notice that 27 of the valid COMP records were missing CT scores. This was due to a room scheduling conflict that arose in the last of the Fall 1988 testing days which resulted in insufficient time to also administer the CT.

Table 9 compares the number of missing CT scores by testing order. There was no significant difference in proportions (P=0.079), between those who took CT first and those who took CT last.

Table 9
Frequencies of Missing and Non-missing CT Scores
For Valid COMP Records Compared by Testing Order

FREQUENCY / PERCENT / ROW PCT / COL PCT	COMP-CT	CT-COMP	TOTAL
NOT MISSING	274 48.24 50.65 96.82	267 47.01 49.35 93.68	541 95.25
MISSING	9 1.58 33.33 3.18	18 3.17 66.67 6.32	27 4.75
TOTAL	283 49.82	285 50.18	568 100.00

CHI-SQUARE (DF=1) = 3.084, P=0.079

Table 10 compares those missing their CT scores with those not missing their CT scores on COMP Total, COMP TOTAL-Gain, ACT and GPA. There were no differences between variances on any of these variables. Furthermore, there were no differences in means for the two COMP-based measures. However those with missing CT scores had significantly lower means on both ACT and GPA.

Table 10
Comparison of Students Missing CT with Those not Missing CT
On Other Relevant Variables For Valid COMP Records

TOTAL CREDIT HOURS EARNED

MISSING	N	MEAN	STD DEV	STD ERROR
NO	541	49.85231054	5.09466489	0.21903675
YES	27	51.54814815	5.04985684	0.97184540

FOR HO: VARIANCES ARE EQUAL, F'(540,26) = 1.02 PROB > F' = 1.0000
FOR HO: MEANS ARE EQUAL T'(566.0) = -1.6887 PROB > |T'| = 0.0918

SOPHOMORE COMP TOTAL SCORE

MISSING	N	MEAN	STD DEV	STD ERROR
NO	541	190.60073937	12.31678768	0.52954006
YES	27	189.22222222	11.07086729	2.13058940

FOR HO: VARIANCES ARE EQUAL, F'(540.26) = 1.24 PROB > F' = 0.5217
FOR HO: MEANS ARE EQUAL T'(566.0) = 0.5701 PROB > |T'| = 0.5688

COMP TOTAL-GAIN SCORE

MISSING	N	MEAN	STD DEV	STD ERROR
NO	541	5.01293900	8.82745186	0.37952180
YES	27	6.88888889	10.65303616	2.05017777

FOR HO: VARIANCES ARE EQUAL, F'(26,540) = 1.46 PROB > F' = 0.1375
FOR HO: MEANS ARE EQUAL T'(566.0) = -1.0666 PROB > |T'| = 0.2866

ENTERING ACT COMPOSITE SCORE

MISSING	N	MEAN	STD DEV	STD ERROR
NO	541	24.39741220	3.79023022	0.16295473
YES	27	22.59259259	3.55462060	0.68408705

FOR HO: VARIANCES ARE EQUAL, F'(540,26) = 1.14 PROB > F' = 0.7211
FOR HO: MEANS ARE EQUAL T'(566.0) = 2.4215 PROB > |T'| = 0.0158

SOPHOMORE CUMULATIVE GPA

MISSING	N	MEAN	STD DEV	STD ERROR
NO	541	2.97046211	0.53424329	0.02296891
YES	27	2.59851852	0.47445275	0.09130847

FOR HO: VARIANCES ARE EQUAL, F'(540,26) = 1.27 PROB > F' = 0.4716
FOR HO: MEANS ARE EQUAL T'(566.0) = 3.5478 PROB > |T'| = 0.0004

This is perhaps because the last sophomore testing day each semester is a "make-up" date for those missing the earlier scheduled day. One might speculate that those who miss their earlier scheduled date tend to be lower ability students.

In summary the resulting data set to be used for analysis in this study contains 568 valid observations. In arriving at this number, 37 students were deleted because they did not show up for their required sophomore examinations and hence had no COMP or CT scores, and 116 were deleted because they had invalid COMP records for one or more reasons. There were no significant differences in measures of freshmen entering ability between those deleted and those retained at either stage of deletion. Predictably GPA was significantly lower for the 37 who did not show up for sophomore examinations. At both stages of deletion there were no significant differences found on any of the other non-COMP variables including sophomore CT performance.

Of the 568 valid COMP records used in this study, 27 students were not administered the CAAP during the Fall 1988 "make-up" date, and consequently these students were eliminated in statistical analyses requiring CT scores. These 27 students did not differ from the rest on the important COMP Total and AP TOTAL-Gain score variables. However, they did differ from the rest on mean entering ACT and cumulative GPA.

The frequencies of deletion and retention were compared by assignment to the two testing orders, CT followed by COMP and COMP followed by CT, at each stage of deletion and for those missing CT scores. In no cases were there differences between CT-COMP and COMP-CT test orders.

Analysis

The analyses to be conducted in this study fall into two logically sequential categories. The first is to test the prerequisite order effect hypothesis. The second, and the target analysis of the study, is to examine the relationship of COMP scores to CT scores making any needed adjustments dependent upon the results of the prerequisite analyses.

The prerequisite hypothesis is that scores do not differ as a result of which test was administered last in a long three and one-half hour testing session -- i.e. order effect. Indeed if fatigue or decreased motivation played a role, CT scores would be lower when CT was given during the last hour than when CT was given during the first hour of the testing session. While the order effect hypothesis is of interest in its own right, it is prerequisite to the target exploratory analyses of relationships between COMP and CT scores, since systematically lower CT scores for half of the sample could obviously distort the relationships under investigation.

Historically NMSU has recognized that scores for those taking sophomore examinations in the spring tend to be systematically higher than the scores for those taking sophomore examinations in the fall. For this reason it seemed natural to control for the semester of sophomore examination in testing the order effect hypothesis. In addition, including semester in the design would allow the testing of possible order by semester interaction effects. Thus a two-way order (CT-COMP and COMP-CT) by semester (Fall 1988 and Spring 1989) analysis of variance design was used to analyze order effects for each of the primary variables of interest in this study--recall Table 2.

The final stage of analysis is the exploration of relationships between the COMP objective test scores and the CAAP CT test scores. This analysis will be done by first focusing on the sophomore COMP process subscores (SP, CV, QCOM, and VCOM) and their associated gain scores defined earlier in Table 2. Then in separate but parallel analyses, the COMP Total score, which is a composite of the four process-oriented subscores, together with COMP TOTAL-Gain, entering ACT composite and cumulative sophomore GPA, will be investigated in relationship to CT. Since freshmen COMP scores are a linear combination of sophomore COMP scores and COMP Gain scores, freshmen COMP scores will not be included in analyses of relationships with CT.

Each set of analyses will include simple zero-order correlations with CT, first-order partial correlations holding ACT constant, and second-order partial correlations holding both ACT and GPA constant. Canonical and partial canonical correlations with CT will also be employed. Additionally, various multiple regression models aimed at predicting CT will be examined to determine the proportion of CT variance that can be accounted for by combinations of variables in the set of process-oriented subscores (SP, CV, VCOM QCOM and their gains) as well as by combinations of variables in the set of composite scores (COMP TOTAL, TOTAL-Gain. ACT and GPA).

At each stage and level of analysis all hypotheses will be tested allowing an alpha=0.05 error rate. All computations will be carried out using Statistical Analysis System (SAS) programing.

RESULTS

Analysis of Test Order Effects

In this study there was a prerequisite hypothesis which needed to be addressed before continuing on to the exploration of COMP and CT relationships. In null form this hypothesis was that there was no difference due to test order effects on any of the test scores to be investigated. Indeed this is an interesting hypothesis in itself due to the potential for student fatigue and motivational factors involved in a three and one-half hour examination session, apart from any psychometric consideration of content or difficulty sequencing. If significant order effects were to be found the meaning of exploratory findings regarding the relationship between the two tests of greatly different lengths (CT=40 minutes, COMP objective test = 150 minutes) would be questionable.

Since sophomore testing takes place in both fall and spring semesters at NMSU, and since at NMSU those sitting for sophomore examinations in the fall historically tend to have lower entering ability as freshmen than those sitting for examinations in the spring, it seemed logical to control for the semester of examination within the design to test for order effects. This also allowed testing the order by semester interaction hypothesis.

Table 11 summarizes the analysis of variance results for the CAAP Critical Thinking (CT), COMP Total, and the freshmen to sophomore COMP TOTAL-Gain scores. On all three of these key variables there were no interactions nor test order effects found. However, consistent with historical findings, both the COMP Total and the COMP TOTAL-Gain scores are greater for those tested in the spring semester.

Table 11
Analysis of Variance Summaries
for Critical Thinking, COMP Total and COMP TOTAL-Gain

DEPENDENT VARIABLE = CAAP Critical Thinking Score

SOURCE	DF	TYPE III SS	F VALUE	PR > F
ORDER	1	1.85204638	0.14	0.7131
SEMESTER	1	30.88213091	2.26	0.1337
ORDER*SEMESTER	1	0.16840705	0.01	0.9117
ERROR	537	7350.29191864		

DEPENDENT VARIABLE = Sophomore COMP Total Score

SOURCE	DF	TYPE III SS	F VALUE	PR > F
ORDER	1	235.21699566	1.58	0.2099
SEMESTER	1	613.90181737	4.11	0.0430
ORDER*SEMESTER	1	288.41463108	1.93	0.1651
ERROR	564	84181.01761325		

DEPENDENT VARIABLE = COMP TOTAL-Gain Score

SOURCE	DF	TYPE III SS	F VALUE	PR > F
ORDER	1	280.03300269	3.57	0.0595
SEMESTER	1	493.15954561	6.28	0.0125
ORDER*SEMESTER	1	199.97455831	2.55	0.1111
ERROR	564	44285.17270696		

To glean a clearer understanding of the semester differences and to check for order by semester interactions, analysis of variance was also conducted on five commonly employed university variables: entering ACT composite score, freshmen COMP Total score, hours of credit earned at the time of sophomore examination, cumulative GPA at the end of sophomore examination year, and student age. Table 12 reports the results of these analyses. As can be seen there were no significant testing order nor order by semester effects on any

Table 12
Analysis of Variance Summaries
for Entering ACT, Freshmen COMP Total, Credit Hours, Cumulative GPA and Age

DEPENDENT VARIABLE - Entering ACT Composite Score

SOURCE	DF	TYPE III SS	F VALUE	PR > F
ORDER	1	8.84583393	0.63	0.4285
SEMESTER	1	211.86881675	15.03	0.0001
ORDER*SEMESTER	1	6.24653322	0.44	0.5058
ERROR	564	7947.91377328		

DEPENDENT VARIABLE - Freshman COMP Total Score

SOURCE	DF	TYPE III SS	F VALUE	PR > F
ORDER	1	1.95272940	0.01	0.9175
SEMESTER	1	2207.51861578	12.13	0.0005
ORDER*SEMESTER	1	8.07434073	0.04	0.8333
ERROR	564	102664.03737580		

DEPENDENT VARIABLE - Total Credit Hours Earned

SOURCE	DF	TYPE III SS	F VALUE	PR > F
ORDER	1	0.74752613	0.04	0.8490
SEMESTER	1	3107.61514843	150.92	0.0001
ORDER*SEMESTER	1	29.07282709	1.41	0.2352
ERROR	564	11613.71081523		

DEPENDENT VARIABLE - Cumulative Grade Point Average

SOURCE	DF	TYPE III SS	F VALUE	PR > F
ORDER	1	0.00184175	0.01	0.9357
SEMESTER	1	3.90457453	13.80	0.0002
ORDER*SEMESTER	1	0.08699048	0.31	0.5794
ERROR	564	159.53445019		

DEPENDENT VARIABLE - Age

SOURCE	DF	TYPE III SS	F VALUE	PR > F
ORDER	1	0.52900257	0.05	0.8223
SEMESTER	1	5.75914675	0.55	0.4588
ORDER*SEMESTER	1	0.58975929	0.06	0.8126
ERROR	564	5909.49250417		

of these five dependent variables. Semester, however, was found to be significant for each dependent variable except student age.

Within the COMP test the process subscores were considered most likely to reflect critical thinking skills as discussed in the introduction. Table 13 reports the order by semester analysis of variance results for these

Table 13
Analysis of Variance Summaries for Sophomore COMP Process Subscores

DEPENDENT VARIABLE - SP (SOLVING PROBLEMS)

SOURCE	DF	TYPE III SS	F VALUE	PR > F
ORDER	1	2.19202605	0.08	0.7730
SEMESTER	1	48.19321947	1.83	0.1766
ORDER*SEMESTER	1	7.40657039	0.28	0.5960
ERROR	564	14847.91214095		26.32608536

DEPENDENT VARIABLE - CV (CLARIFYING VALUES)

SOURCE	DF	TYPE III SS	F VALUE	PR > F
ORDER	1	34.41777236	1.69	0.1944
SEMESTER	1	16.25584931	0.80	0.3723
ORDER*SEMESTER	1	29.41996319	1.44	0.2302
ERROR	564	11501.70374205		20.39309174

DEPENDENT VARIABLE - VCOM (VERBAL COMMUNICATION)

SOURCE	DF	TYPE III SS	F VALUE	PR > F
ORDER	1	21.53190124	2.29	0.1311
SEMESTER	1	23.99680994	2.55	0.1110
ORDER*SEMESTER	1	12.54625416	1.33	0.2490
ERROR	564	5313.26709761		9.42068634

DEPENDENT VARIABLE - QCOM (QUANTITATIVE COMMUNICATION)

SOURCE	DF	TYPE III SS	F VALUE	PR > F
ORDER	1	2.94640951	0.20	0.6568
SEMESTER	1	43.32369236	2.91	0.0888
ORDER*SEMESTER	1	16.16924457	1.08	0.2981
ERROR	564	8406.84612850		14.90575555

COMP process-oriented subscores. As seen from this table, there were no order effects, nor were any other effects significant.

Table 14 summarizes analysis of variance findings for the four process-oriented freshmen to sophomore gain scores defined from the COMP

Table 14
Analysis of Variance Summaries
for COMP Process Subscore Gains

DEPENDENT VARIABLE = SP-GAIN

SOURCE	DF	TYPE III SS	F VALUE	PR > F
ORDER	1	0.56749487	0.02	0.8940
SEMESTER	1	153.90639173	4.82	0.0285
ORDER*SEMESTER	1	3.97101116	0.12	0.7244
ERROR	564	18002.90913516		31.92005166

DEPENDENT VARIABLE = CV-GAIN

SOURCE	DF	TYPE III SS	F VALUE	PR > F
ORDER	1	9.96003680	0.38	0.5368
SEMESTER	1	82.47038525	3.16	0.0759
ORDER*SEMESTER	1	6.00885408	0.23	0.6314
ERROR	564	14706.50134622		26.07535700

DEPENDENT VARIABLE = VCOM-GAIN

SOURCE	DF	TYPE III SS	F VALUE	PR > F
ORDER	1	10.80999941	0.47	0.4942
SEMESTER	1	4.47556519	0.19	0.6600
ORDER*SEMESTER	1	3.73203280	0.16	0.6879
ERROR	557	12869.10822918		23.10432357

DEPENDENT VARIABLE = QCOM-GAIN

SOURCE	DF	TYPE III SS	F VALUE	PR > F
ORDER	1	31.32431491	2.29	0.1306
SEMESTER	1	8.16059850	0.60	0.4400
ORDER*SEMESTER	1	13.63712857	1.00	0.3183
ERROR	557	7612.17521713		13.66638280

objective test. One significant difference was found on the problem solving gain score for the spring semester (Mean = 0.4377) as compared to the fall semester (Mean = 1.5975) of sophomore examination, which was consistent with the direction of differences found between semesters on the COMP TOTAL-Gain score. Otherwise Table 14 shows no evidence of effects due to test order or to order by semester interaction.

Table 15 reports the means for each dependent variable in the preceding analysis of variance tables by semester. Notice that those taking sophomore examinations in the spring had significantly higher entering levels of ability as seen from both the ACT and the freshmen COMP Total score. It is therefore

Table 15
Mean Scores by Semester

	FALL 1988			SPRING 1989			
	N	MEAN	STD DEV	N	MEAN	STD DEV	P<0.05
CRIT. THINKING	132	66.3258	3.8214	409	66.8875	3.6518	
COMP TOTAL	159	188.8742	11.8347	409	191.1809	12.3689	Yes
COMP TOTAL-GAIN	159	6.6038	9.9626	409	4.5183	8.4221	Yes
ENTERING ACT	159	23.3333	3.7697	409	24.6919	3.7420	Yes
FRES COMP TOTAL	159	182.2704	13.0295	409	186.6626	13.6354	Yes
EARNED CREDITS	159	53.6843	5.6848	409	48.4746	4.0036	Yes
CUMULATIVE GPA	159	2.8198	0.5498	409	3.0045	0.5237	Yes
AGE	159	19.6981	3.0580	409	19.4743	3.2966	
SP	159	76.7799	4.6192	409	77.4279	5.3054	
CV	159	57.1132	4.8834	409	57.4866	4.3662	
VCOM	159	32.2076	3.3922	409	32.6626	2.9370	
QCOM	159	22.3711	3.5535	409	22.9853	3.9692	
SP GAIN	159	1.5975	5.5713	409	0.4377	5.6670	Yes
CV GAIN	159	1.3396	5.5116	409	0.4890	4.9304	
VCOM GAIN	155	1.1935	4.8779	406	0.9951	4.7702	
QCOM GAIN	155	1.9742	3.7520	406	2.2463	3.6775	

not surprising that the spring-tested group also had the higher GPA. Also consistent is the fact that the number of credit hours earned at the time of sophomore examination was greater for fall-tested students, since those students are more likely to be those who earn fewer credits each semester; and, thus must wait until fall to have over the 44 credit hours needed to qualify for sophomore examinations. Indeed 98.0% of those tested in the spring of 1989 entered the university three semesters earlier, in the fall of 1987, whereas, 30.6% of those tested in the fall of 1988 entered in the fall of 1987. In fact 60.5% of those taking sophomore examinations in the fall of 1988 entered in the fall of 1986 (four semesters earlier) with 8.9% who entered earlier than 1986. Clearly the fall examinees tend to have different characteristics than do the mainstream of students who sit for sophomore examinations in the spring--a pattern that seems to be present every academic year.

Based upon the evidence provided in this section, one may conclude that the order in which students took the COMP and CT tests has no significant impact upon the score results. This finding is of interest in itself since it demonstrates that even if fatigue or decreases in motivation occur from the beginning to the end of a lengthy (3-4 hours) required testing session, such fatigue or loss of motivation (if it exists at all) does <u>not</u> significantly reflect itself in student performance on these examinations. Those who took the CT test immediately before the COMP test (CT Mean = 66.839) did not differ from those who took the CT after the COMP (CT Mean = 66.664). Finally for the purposes of this study, the order effect results reported here allows progression to the exploration of CT and COMP relationships in a much cleaner and more interpretable fashion.

Analysis of CT and COMP Relationships

The exploration of relationships between CT and COMP was done using partial correlations and multiple regression models for predicting CT. First the analysis of the relationship of CT scores to the process-oriented subscores of the COMP and their associated gain scores was conducted. Secondly partial correlations and regression models were used to look at COMP Total scores in relationship to CT. As part of this second set of analyses ACT and GPA were also examined as potential predictors of CT. Since six students had missing COMP computation subscores (QCOM), the analyses involving the COMP subscores were conducted on a sample of 535 students, whereas, the other analyses were based upon all 541 students who took the CAAP Critical Thinking test.

Table 16 gives descriptive statistics for the sophomore COMP process subscores and their related gain scores together with the correlations of these COMP-based scores with CT, ACT and GPA. The bottom portion of Table 16 gives the intercorrelations among CT, ACT and GPA, as well as their correlations with COMP Total and TOTAL-Gain. Notice that all of the COMP status scores (SP, CV, QCOM and VCOM) have significant positive correlations with CT, ACT and GPA. Furthermore, eight out of twelve correlations are significant for COMP gain scores, and of those eight, only the three with QCOM-Gain are positive. None of the TOTAL-Gain correlations are significant.

In looking at the correlations with CT, it seems evident that SP, CV, QCOM and VCOM are related to critical thinking. However, CT correlations with COMP freshmen to sophomore gain scores are not so consistent. The correlation of CT with CV-Gain is not significantly different from zero, and the small CT with QCOM-Gain correlation is the only significant, positive

Table 16
Descriptive Statistics and Correlations with CT, ACT and GPA

	PEARSON CORRELATIONS PROB > |R| UNDER HO:RHO=0			UNIVARIATE STATISTICS		
	CT	GPA	ACT	N	MEAN	STD
SP SOLVING PROBLEMS	0.41954 0.0001	0.30361 0.0001	0.50861 0.0001	535	77.2561	5.14580
CV CLARIFYING VALUES	0.44630 0.0001	0.31804 0.0001	0.53452 0.0001	535	57.3514	4.51214
QCOM QUANTITATIVE COMMUNICATION	0.43125 0.0001	0.40340 0.0001	0.60295 0.0001	535	22.8505	3.90382
VCOM VERBAL COMMUNICATION	0.34561 0.0001	0.22041 0.0001	0.34087 0.0001	535	32.5383	3.05228
SP-GAIN SOLVING PROBLEMS GAIN	-0.08917 0.0392	0.01915 0.6585	-0.03842 0.3751	535	0.6935	5.66475
CV-GAIN CLARIFYING VALUES GAIN	-0.06364 0.1415	-0.09977 0.0210	-0.05145 0.2348	535	0.6579	5.13970
QCOM-GAIN QUANTITATIVE COMM. GAIN	0.19223 0.0001	0.13805 0.0014	0.18334 0.0001	535	2.2019	3.71408
VCOM-GAIN VERBAL COMM. GAIN	-0.18485 0.0001	-0.23093 0.0001	-0.29057 0.0001	535	1.0393	4.78335
CT CRITICAL THINKING	1.00000 0.0000	0.42776 0.0001	0.62070 0.0001	535	66.7514	3.69531
GPA CUMULATIVE GPA	0.42776 0.0001	1.00000 0.0000	0.54510 0.0001	535	2.9704	0.53299
ACT ENTERING ACT	0.62070 0.0001	0.54510 0.0001	1.00000 0.0000	535	24.4150	3.76806
COMP TOTAL	0.60127 0.0001	0.46193 0.0001	0.73286 0.0001	535	190.5626	12.32197
COMP TOTAL-GAIN	-0.05539 0.2009	-0.05873 0.1750	-0.06262 0.1481	535	4.9365	8.76918

relationship (r=0.192). While the CT correlations with SP-Gain and VCOM-Gain are significant, both are negative, indicating that higher freshmen to sophomore gains tend to go with lower levels of critical thinking. It is interesting that the two portions of the usual COMP communications gain score, given in Table 16 as QCOM-Gain and VCOM-Gain, are significantly correlated with CT in different directions, QCOM-Gain being positive and VCOM-Gain being negative. Based upon these data, there is very little evidence that higher gains on these COMP process variables might be associated with higher critical thinking. Indeed only the QCOM-Gain had a significant and positive relationship to CT, while the SP-Gain and the VCOM-Gain had significant negative relationships with CT, and the CV-Gain correlation with CT was not significantly different from zero. Notice that subscore gains tend to cancel each other out to leave a nonsignificant TOTAL-Gain with CT correlation.

Table 16 shows that the highest correlation observed was between COMP Total and ACT. In reviewing the correlations of COMP-based variables with ACT and GPA, Table 16 further shows that the SP, CV and QCOM correlations with ACT are higher than their correlations with CT, while their correlations with GPA are lower than with CT though still clearly significant. The VCOM correlation with ACT, however, does not significantly differ from its correlation with CT.

For the COMP-based gain scores, SP-Gain is not significantly related to either ACT or GPA. For GPA, a small negative correlation with CV-Gain (r=-0.0998) exists, although CV-Gain is not significantly related to either ACT or CT. For QCOM-Gain and VCOM-Gain, the same pattern of correlations exist with both ACT and GPA as existed with CT--i.e. positive correlations with QCOM-Gain and negative with VCOM-Gain. The negative correlations with gains would indicate that those with lower entering ability and those with lower levels of achievement during the first 3-4 semester of college tend to

make the greatest gains on the COMP process measures. While there may be several explanations for the small negative yet significant gain score correlations with CT, ACT and GPA (e.g. programs with greater educational impact for lower ability students, testing and gain score artifacts, chance variation, possibly greater motivation for low ACT and GPA students, regression to the mean, etc.), for the purposes of this study it seems sufficient to recognize the leverage that ACT and GPA might have on CT and COMP relationships. For this reason it would seem reasonable to examine partial correlations where ACT and GPA are held constant.

Table 17 presents zero-order, first-order and second-order partial correlations controlling for both ACT and GPA as well as canonical correlations of COMP-based sets of scores with CT. As can be seen from Table 17, partialling out entering level ability, as measured by the ACT, lowers COMP-score with CT correlations to a greater extent than does partialling out GPA as a measure of achievement over the first 3-4 semesters of college. Furthermore, there seem to be no appreciable differences between first-order partial correlations controlling for ACT and second-order correlations controlling for both ACT and GPA.

In comparing zero-order correlations with first-order correlations controlling for ACT, Table 17 shows that holding ACT constant did not lower the VCOM correlation as much as the SP, CV and QCOM correlations. Indeed with either ACT or both ACT and GPA held constant, VCOM has the greatest correlation with CT, whereas, its zero-order correlation with CT was the smallest. Also, when one looks at the gain score correlations, one recognizes that the partialling out of ACT left a nonsignificant VCOM-Gain with CT

Table 17
Zero- First- and Second-Order Partial Correlations
Between CT with COMP Subscores Controlling for ACT and GPA

COMP VARIABLE	Partial Correlations with CT (P for H0: r=0) Controlling for: 0	ACT	GPA	ACT & GPA
SP	0.41954 (0.0001)	0.153832 (0.0004)	0.336340 (0.0001)	0.150360 (0.0005)
CV	0.44630 (0.0001)	0.172823 (0.0001)	0.362039 (0.0001)	0.169396 (0.0001)
QCOM	0.43125 (0.0001)	0.091135 (0.0353)	0.312782 (0.0001)	0.077131 (0.0752)
VCOM	0.34561 (0.0001)	0.181833 (0.0001)	0.285055 (0.0001)	0.177683 (0.0001)
SP-GAIN	-0.08917 (0.0392)	-0.083375 (0.0542)	-0.107734 (0.0127)	-0.090833 (0.0360)
CV-GAIN	-0.06364 (0.1415)	-0.040499 (0.3503)	-0.023313 (0.5909)	-0.029224 (0.5008)
QCOM-GAIN	0.19223 (0.0001)	0.101755 (0.0187)	0.148764 (0.0006)	0.096464 (0.0259)
VCOM-GAIN	-0.18485 (0.0001)	-0.005985 (0.8903)	-0.097864 (0.0237)	-0.006400 (0.8828)
Canonical { SP CV QCOM VCOM }	0.602048 (0.0001)	0.276344 (0.0001)	0.503744 (0.0001)	0.267292 (0.0001)
Canonical { 4 GAINS }	0.269809 (0.0001)	0.131810 (0.0543)	0.199719 (0.0002)	0.130428 (0.0592)
Canonical { ALL 8 }	0.631374 (0.0001)	0.332543 (0.0001)	0.540154 (0.0001)	0.326051 (0.0001)

correlation of -0.006, whereas, its significant zero-order correlation with CT was -0.185.

With the possible exception of the gain score correlations just noted, the partial correlations with gain scores reported in Table 17 generally offer very little in the way of interpretable information regarding the relationship between CT and COMP. Indeed, the reader should be reminded at this point that, by their nature, gain score zero-order correlations with CT are already similar to partial correlations (partialling out the highly correlated freshmen version of the score leaving a residual from the sophomore score to correlate with something else--recall that Gain = Sophomore - Freshmen).

Table 17 additionally illustrates that the greatest relative reduction in correlation as a result of partialling out ACT, takes place for the QCOM with CT correlation. Moreover when the effects of both ACT and GPA are controlled, the correlation of QCOM with CT is no longer significant (P=0.077). It is again interesting that the QCOM and VCOM portions of the usual communications COMP subscore behave in opposite ways relative to CT.

The canonical correlations with CT are reported at the bottom of Table 17. For the set of SP, CV, QCOM and VCOM the zero-order canonical correlation with CT is 0.602, which is clearly significant. For the set of all eight COMP-based variables in the upper portion of Table 17, the canonical correlation with CT is 0.631, which indicates that inclusion of the additional gain score variables does not contribute much (0.03) to the relationship with CT. This difference is slightly greater as one controls for ACT, GPA or both, yet the same basic pattern is in place. While the partial canonical correlations are all significant for the status set of four COMP subscores and for the set of all eight COMP-based scores, the canonical partial correlations

of CT with the set of four gain scores are not significant when controlling either for ACT or for both ACT and GPA (alpha=0.05).

What can one conclude from the correlations just described? First, there are clearly significant zero-order correlations between COMP process-oriented subscores and CT, both separately and in aggregate. However, these correlations could be due to a strongly related third set of variables with which both CT and COMP are correlated. Two commonly used university variables that should logically be correlated with both CT and COMP are entering ACT composite and cumulative GPA; indeed, Table 1 proved this to be the case. Partialling out ACT to hold entering level of ability/achievement constant reduced the CT with COMP correlations, yet these resulting correlations were still significant. Similarly, partialling out GPA to hold achievement level attained since entry to college constant, left reduced yet significant CT with COMP correlations. In short by controlling for these two important variables underlying both CT and COMP, significant CT correlations still exist with SP, CV, VCOM and QCOM, both separately and as a set (canonical R=0.267). So even in a somewhat purified sense the conclusion is that CT and COMP process subscores are related.

On the other hand, referring to the CT with COMP partial correlations as "somewhat pure" is perhaps an impractical notion. After all critical thinking presumes some ability to think and some background level of achievement to provide something to think about. Thus attempting to partial out ACT and GPA or similar variables may be a case of over control.

In terms of the amount of useful information a college might obtain about critical thinking from COMP, one would not consider the relationships that have been presented strong enough to warrant using COMP variables in place of CT if one was truly interested in obtaining a rather precise estimate

of critical thinking as defined by CAAP. Of course one might speculate that there may be a reasonable alternative definition of critical thinking that could be better estimated by COMP or a combination of COMP and CAAP, but such a definition would need to be tied to more specific components of the tests (e.g. specific items) than is possible at the level of the current analysis. On the other hand, regression estimates of CT using COMP variables as predictors might be of some practical usage in obtaining loose estimates of CT for programs, without the time and expense of administering the additional test. This leads to the next stage of exploring the relationship of CT with COMP -- namely regression models using CT as the criterion variable.

Table 18 summarizes the results of the stepwise procedure for selecting a regression model where all eight COMP-based variables (SP, CV, QCOM, VCOM, SP-Gain, CV-Gain, QCOM-Gain and VCOM-Gain) were available for inclusion. Both the significance levels for entry and for staying in the model were set at 0.15. The most potent predictor was CV which accounted for nearly 20% of the

Table 18
Stepwise Regression Results
Using Eight COMP Subscores as Potential Predictors of CT

STEP	VARIABLE ENTERED	VARIABLE REMOVED	NUMBER IN	PARTIAL R**2	MODEL R**2	F	PROB>F
1	CV		1	0.1992	0.1992	132.5688	0.0001
2	QCOM		2	0.0867	0.2859	64.5781	0.0001
3	SP		3	0.0459	0.3318	36.4763	0.0001
4	SP-GAIN		4	0.0405	0.3723	34.1881	0.0001
5	VCOM		5	0.0160	0.3882	13.7998	0.0002
6	VCOM-GAIN		6	0.0057	0.3940	5.0007	0.0258
7	CV-GAIN		7	0.0047	0.3986	4.0927	0.0436

CT variance. Notice that only the QCOM-Gain did not meet criteria for entry, after the other seven variables had been included. It is also interesting that SP-Gain was included before VCOM. In total about 40% of the variance is accounted for using these seven COMP-based scores as predictors of CT, however for practical reasons it is worth noting that entry of the last two variables only accounted for about a 1% addition.

Table 19 presents, in order of magnitude, the R-square values for selected CT prediction models. For each "number of variables in model" category, the first R-square value is the smallest obtained for all possible combinations of that number of variables, whereas, the last R-square value in the category is the largest obtained for all possible combinations of that number of variables. Notice that using all eight predictors accounts for virtually no additional variance compared to the seven variable model found by the stepwise procedure. It is also of practical interest to recognize that the four variable model using SP, CV, QCOM and VCOM accounts for 36.3% of the variance, only 1% less variance than the best (by R-square magnitude) four variable model which gives preference to SP-Gain over VCOM. Indeed, SP-Gain in combination with SP and QCOM presents the best three variable model accounting for 34.0% of the variance, whereas, the best two variable model, CV and QCOM, accounts for 28.6% of the CT variance. Excluding gain scores from consideration, SP, VC and QCOM as predictors of CT account for 33.2% of the variance. In summary it seems clear that the COMP-based process subscores can account for at least one third of the variance in critical thinking, and up to 40% with COMP-based gain scores included.

While the focus has been on the COMP subscores most likely to be related to critical thinking, it is also worth examining the relationship of the COMP Total and COMP TOTAL-Gain to CT since these scores are commonly available and

Table 19
R-square Values for Selected
Regression Models to Predict CT from COMP Subscores

NO. IN MODEL	R-SQUARE (N=535)	VARIABLES IN MODEL
2	0.0120483	SP-GAIN CV-GAIN
2	0.0617106	QCOM-GAIN VCOM-GAIN
2	0.2711506	QCOM VCOM
2	0.2725078	SP CV
2	0.2803958	SP SP-GAIN
2	0.2858678	CV QCOM
3	0.0447984	SP-GAIN CV-GAIN QCOM-GAIN
3	0.3022066	SP CV VCOM
3	0.3295640	CV QCOM VCOM
3	0.3308935	SP CV SP-GAIN
3	0.3317708	SP CV QCOM
3	0.3403987	SP QCOM SP-GAIN
4	0.0727968	SP-GAIN CV-GAIN QCOM-GAIN VCOM-GAIN
4	0.2990724	SP SP-GAIN QCOM-GAIN VCOM-GAIN
4	0.3485843	SP CV QCOM CV-GAIN
4	0.3625288	SP CV QCOM VCOM
4	0.3644284	SP QCOM VCOM SP-GAIN
4	0.3722635	SP CV QCOM SP-GAIN
5	0.2090088	QCOM SP-GAIN CV-GAIN QCOM-GAIN VCOM-GAIN
5	0.2992031	SP SP-GAIN CV-GAIN QCOM-GAIN VCOM-GAIN
5	0.3707012	SP CV QCOM VCOM VCOM-GAIN
5	0.3744734	SP QCOM VCOM SP-GAIN VCOM-GAIN
5	0.3804507	SP CV QCOM SP-GAIN CV-GAIN
5	0.3882226	SP CV QCOM VCOM SP-GAIN
6	0.3079465	QCOM VCOM SP-GAIN CV-GAIN QCOM-GAIN VCOM-GAIN
6	0.3599947	SP CV SP-GAIN CV-GAIN QCOM-GAIN VCOM-GAIN
6	0.3810374	SP CV QCOM SP-GAIN CV-GAIN QCOM-GAIN
6	0.3882906	SP CV QCOM VCOM SP-GAIN QCOM-GAIN
6	0.3932845	SP CV QCOM VCOM SP-GAIN CV-GAIN
6	0.3939625	SP CV QCOM VCOM SP-GAIN VCOM-GAIN
7	0.3629843	CV QCOM VCOM SP-GAIN CV-GAIN QCOM-GAIN VCOM-GAIN
7	0.3751603	SP QCOM VCOM SP-GAIN CV-GAIN QCOM-GAIN VCOM-GAIN
7	0.3790040	SP CV QCOM VCOM CV-GAIN QCOM-GAIN VCOM-GAIN
7	0.3933476	SP CV QCOM VCOM SP-GAIN CV-GAIN QCOM-GAIN
7	0.3939628	SP CV QCOM VCOM SP-GAIN QCOM-GAIN VCOM-GAIN
7	0.3986327	SP CV QCOM VCOM SP-GAIN CV-GAIN VCOM-GAIN
8	0.3986330	SP CV QCOM VCOM SP-GAIN CV-GAIN QCOM-GAIN VCOM-GAIN

since they are a composite of the subscores just explored. Additionally, if prediction of CT is actually desired, ACT and GPA are worth considering as potentially potent predictors.

Table 20 provides the zero-order, first-order and second-order partial correlations with CT for the more global variables, COMP Total, TOTAL-Gain, ACT and GPA. Notice that only the TOTAL-Gain score correlations with CT were not significant. Comparing the partial correlations of COMP Total and CT in Table 20 with the canonical correlations of the set of process subscores (SP, CV, QCOM & VCOM) with CT in Table 17, one recognizes that these correlations are very similar, as they should be since the COMP Total is a composite of these four subscores. The COMP Total score offers about the same power to predict CT as the four process scores examined earlier; and, prediction based only on COMP Total would be simpler.

Table 20
Zero- First- and Second-Order Partial Correlations
of CT with COMP TOTAL and COMP TOTAL-Gain, Controlling for ACT and GPA

	Partial Correlations with CT (P for HO: r=0)			
	Controlling for:			
COMP VARIABLE	0	ACT	GPA	ACT & GPA
COMP TOTAL	0.60138 (0.0001)	0.27240 (0.0001)	0.50320 (0.0001)	0.26173 (0.0001)
TOTAL-GAIN	-0.06230 (0.1479)	-0.02331 (0.5885)	-0.03654 (0.3967)	-0.01871 (0.6648)
ACT	0.62433 (0.0001)		0.51493 (0.0001)	
GPA	0.42987 (0.0001)	0.13522 (0.0016)		

Table 21 summarizes the stepwise regression results using COMP Total, COMP TOTAL-Gain, ACT and GPA as the potential predictors of CT. All four variables were selected as significant contributors to the model, though the addition of the last variable selected, GPA, accounts only for a little more than 0.5% additional CT variance. It is interesting to note that with ACT and COMP TOTAL already in the model, TOTAL-Gain accounts for more additional variance than does GPA.

Table 21
Stepwise Regression Results
Using COMP TOTAL, TOTAL-GAIN, ACT and GPA as Potential Predictors of CT

STEP	VARIABLE ENTERED	VARIABLE REMOVED	NUMBER IN	PARTIAL R**2	MODEL R**2	F	PROB>F
1	ACT		1	0.3898	0.3898	344.3008	0.0001
2	COMP TOTAL		2	0.0453	0.4351	43.1199	0.0001
3	TOTAL-GAIN		3	0.0106	0.4457	10.2843	0.0014
4	GPA		4	0.0056	0.4513	5.5026	0.0194

Table 22 documents the R-square values for all possible regression models involving combinations of the four potential predictors of CT given in Table 21. In comparing these results to those given earlier in Table 19, notice that COMP TOTAL together with COMP TOTAL-Gain accounts for only 0.3% less CT variance than did all eight of the variables in Table 19. Also one may wish to note that without the gain score TOTAL, ACT and GPA account for only about 1% less CT variance than all four variables.

Table 22
R-square Values for Selected
Regression Models to Predict CT from COMP Total Scores

NUMBER IN MODEL	R-SQUARE (N = 541)	VARIABLES IN MODEL
1	0.00388125	TOTAL-GAIN
1	0.18478627	GPA
1	0.36166137	TOTAL
1	0.38978885	ACT
2	0.18587497	TOTAL-GAIN GPA
2	0.39012101	TOTAL-GAIN ACT
2	0.39120781	TOTAL GPA
2	0.39568610	TOTAL TOTAL-GAIN
2	0.40094564	ACT GPA
2	0.43506740	TOTAL ACT
3	0.40115524	TOTAL-GAIN ACT GPA
3	0.41542522	TOTAL TOTAL-GAIN GPA
3	0.44198150	TOTAL ACT GPA
3	0.44568331	TOTAL TOTAL-GAIN ACT
4	0.45131614	TOTAL TOTAL-GAIN ACT GPA

It is important to point out that the ACT, which was given before entry to college, is a better predictor (R-square=0.390) of CT by itself than the COMP given at the same time as the CAAP CT test during sophomore examinations. Moreover, as seen from both Tables 21 and 22, the addition of COMP Total to ACT only accounts for an additional 4.5% of the CT variance. Indeed, it can be estimated from Table 16 that ACT accounts for 53.7% of the COMP Total score variance. Consequently, the evidence indicates that ACT may actually be slightly more predictive of CT than COMP.

Table 23 summarizes relevant regression model parameter estimates for predicting CT scaled scores. Should readers be inclined to utilize one of these CT prediction equations, they are cautioned to carefully review the definitions of variables found in Table 1 as well as the characteristics of the sample from which these estimates were obtained.

Table 23
Regression Model Parameter Estimates for Predicting CT Score

	PARAMETER ESTIMATES					
INTERCEPT	29.95183	31.98870	34.70378			
SP	0.20134	0.14637	0.20871			
CV	0.18154	0.18137	0.27753			
QCOM	0.18755	0.25182				
VCOM	0.20711	0.22407				
SP GAIN	-0.11616					
CV GAIN	-0.06038					
VCOM GAIN	-0.06807					
INTERCEPT	32.33404	51.88835	30.60677	39.47506	37.44630	39.24789
COMP TOTAL	0.18057		0.19170	0.09355	0.10668	0.08956
ACT		0.60917		0.38709	0.30047	0.34320
COMP TOTAL-GAIN			-0.07882		-0.04373	
GPA					0.62625	0.69155

CONCLUSIONS

This study was aimed at the exploration of the relationship between COMP test scores and CAAP Critical Thinking test scores. Total semester hours of credit and entering ACT scores were used in the stratified design to assign sophomore tests to students. However to address the possibility of fatigue, loss of motivation and any test order effects within an approximately three and one-half hour sophomore testing session, a prerequisite hypothesis tested was that there were no differences between the scores of those who were administered the CT first and those who were administered the CT last.

The results of the sub-study of test order effects showed that there were no significant (alpha=0.05) differences between the scores obtained from the two testing orders, COMP-CT and CT-COMP. Furthermore there were no significant interactions of testing order with the controlling variable SEMESTER (Fall & Spring). It was therefore concluded that effects did not exist due to fatigue, losses of motivation over the time span of the testing session and/or the sequence of tested content; or, at least, if they did exist they did not alter test performance in any significant way. This finding was of some importance in itself since it counters the argument that students will perform less well near the end of a long testing session than near the beginning, and therefore it has implications for campus assessment programs. However, for the primary purpose of this study, the order effect results allowed progression to the exploration of relationships between CT and COMP without problematical adjustments or indirect interpretations being required.

The relationship between COMP and CT was explored by first focusing on four process-oriented COMP subscores (SP, CV, QCOM and VCOM) and their related freshman to sophomore gain scores. Recognizing high correlations between ACT and these subscores, as well as lower yet, significant correlations between

GPA and these COMP process subscores, partial correlations holding ACT and GPA constant were examined. The same zero-order, first-order, and second-order partial correlations were also examined for the four gain scores.

With both ACT and GPA partialled out, SP, CV and VCOM had significant correlations ranging between 0.15 for SP and 0.18 for VCOM, whereas, the four zero-order correlations ranged from 0.35 for VCOM to 0.45 for CV. The CT with QCOM correlation was no longer significant after holding ACT and GPA constant. The canonical partial correlation between CT and the set of these four variables was 0.27. Thus with entering level of ability (ACT) and level of achievement during the first 3-4 semesters of college (GPA) held constant, small yet significant second-order partial correlations between COMP process subscores and CT scores were present.

The correlations of CT with the four process related gain scores were substantially less interpretable. The CV-Gain with CT zero-order correlation was not significant. The zero-order correlations with SP-Gain and VCOM-Gain were significant and negative (-0.089 and -0.182 respectively), but after ACT and GPA were partialled out, only SP-Gain had a significant second-order partial correlation with CT (-0.091). The computing part of the usual COMP communications subscore, defined as QCOM in this study, had the only positive zero-order gain score correlation with CT (0.192), and this correlation remained significant after ACT and GPA were partialled out (0.096). Thus given the mixed set of gain score correlations with CT, there is insufficient evidence to support the contention that higher gains are associated with higher levels of critical thinking.

In considering the correlations with gain scores, the nature of gain scores should be taken into account. Since gain scores have been defined as sophomore performance minus freshmen performance, they are partial scores.

Thus zero-order correlations with gain scores are similar to first-order partial correlations controlling for freshmen performance, and first-order partial correlations with gain scores are similar to second-order, etc.

It was observed that the verbal communication (VCOM) and the quantitative communication (QCOM) portions of the usual communication subscore reported by COMP behaved in almost opposite ways relative to CT. VCOM had the lowest zero-order correlation with CT, yet after partialling out ACT and GPA it had the highest correlation with CT. QCOM on the other hand had a high zero-order correlation with CT, but the second-order partial correlation with CT was the lowest and did not even meet significance criteria. Also when looking at gain score correlations, the zero-order correlations with CT were 0.192 for QCOM-Gain and -0.185 for VCOM-Gain.

The relationships between CT and COMP-based subscores and gains were explored further using multiple regression with CT as the criterion variable. Employing stepwise regression procedures to identify the "best" significant model, all four subscores and three of the gain scores were selected, which accounted for 39.9% of the CT variance. The score excluded by this procedure was QCOM-Gain. Excluding all gain scores and using the four subscores SP, CV, QCOM and VCOM accounted for 36.3% of the CT variance.

The relationships of CT to the COMP Total score, the TOTAL-Gain score, ACT and GPA were also explored using both partial correlations and multiple regression methods. Of these scores ACT had the highest zero-order correlation with CT (0.624), whereas, the COMP Total correlation with CT was slightly less (0.601). The correlation of TOTAL-Gain with CT was not significant. With both ACT and GPA held constant the second-order partial correlation of COMP Total with ACT was a significant 0.262.

The multiple regression model using COMP Total, TOTAL-Gain, ACT and GPA as predictors of CT accounted for 45.1% of the variance, whereas, excluding the gain scores left 44.2% of the variance accounted for. It was observed that ACT alone accounted for more CT variance (39.0%) than did COMP Total alone (36.2%) -- ACT and COMP Total share 53.7% of their variances in common. In light of the fact that ACT is typically administered at least two years earlier than the sophomore examinations in which both COMP and CT were administered, it was a little surprising that ACT maintained a greater relationship to CT than did COMP which was administered at the same time and which seems to offer more tasks requiring critical thinking skills.

In reviewing the exploratory relationships just summarized, it is important to recognize that critical thinking assumes, as a prerequisite, the ability to think and some content knowledge providing something to think about. Thus attempting to partial out entering ability via ACT or college content achievement via GPA may leave conservative estimates of relationship in the practical sense of what critical thinking involves. In this vein of reasoning, the high correlation between CT and ACT is not so surprising. Indeed, looking at it from a different angle, ACT is designed to assess the ability of students to handle (think with respect to) higher (postsecondary) levels of basic content (achievement domains). In other words, the ability/achievement (A) prerequisite to college level thinking which implies critical thinking (CT), gives a new conceptual meaning to the acronym "ACT" (A+CT), which not only accounts (in part) for the success of ACT as a predictive measure for admission purposes, but in the context of this study, aids in a more practical understanding of the relationships observed.

A related caution has to do with the relationship between COMP and ACT. While arguments similar to those just given regarding CT and ACT may hold here

as well, it should be recalled that one of the screening procedures for detecting and eliminating invalid COMP scores utilized ACT to COMP concordance tables. While this procedure is highly desirable, it will tend to slightly enhance the ACT with COMP relationship in comparison to relationships between other scores that did not benefit from this screening (e.g. ACT with CT).

It may also be worth noting that much of the ACT test is similar in design to the CAAP-CT with passages followed by questions, whereas, the COMP has a much wider range of activities and presentation. It is possible that the ACT with CT correlation may be slightly enhanced in comparison to COMP with CT correlation due to this greater similarity in format and presentation.

An additional consideration in the interpretation of the correlational evidence provided in this study is that statistics are based upon students at a rather selective institution (entering mean ACT = 24.4 for those taking sophomore examinations). For students with ACT scores over 28 a slight ceiling effect may exist on the CT, which could have the effect of slightly depressing correlations with this variable. Also the selective nature of the data would tend to enhance the degree of distributional truncation that typically occurs through the process of college admissions. This would also tend to have a diminishing effect on correlational estimates.

Parameter estimates for regression equations have been reported. Should readers wish to use one of these within their own assessment programs, they are cautioned to examine the data-base characteristics and the definitions of the scores employed in this study. Furthermore, the emphasis in this study is on program assessment rather than individual assessment. It is <u>not</u> suggested that these prediction models would be appropriate for reliably estimating individual student CT scores.

Finally the distinction between "measurement" and "prediction" must be considered. Measurement attempts to be directly representative of the underlying attribute by systematically observing indicators of the attribute's presence and functionally mapping observed indications onto a scale. Prediction only necessitates that a predictor attribute (e.g. human weight) be related to the criterion attribute (e.g. human height). A measure of one is not a substitute measure for the other, even though some information about one attribute can be obtained by recognizing a relationship. This would seem to be the case in this study. That is, while there are clear and significant relationships between COMP and CT that allow various levels of prediction, and while there are undoubtedly some items in the COMP (or for that matter the ACT) that might be usable on a CT-like measure, there is insufficient evidence to conclude that COMP measures the same thing that CT measures. Neither COMP nor ACT (nor both combined) is a substitute measure for CT, if what CT measures is what one is after. However if one wished to obtain a loose estimate of CT for a program group based upon COMP and other related scores, that may be possible with the cautions advised.

A logical and perhaps badly needed extension to the exploratory study reported here would be to examine these sets of relationships within (or controlling for in other ways) specific college level content areas. After all, critical thinking does not typically happen in a content-free fashion; one must have something to critically think about. Indeed one might even hypothesis that relationships between CT and COMP sub-areas would be different for different university programs and different patterns of course work. The identification of such relationships could aid in the development or enhancement of programs to further benefit the education of our students.

REFERENCES

ACT, Report on the Technical Characteristics of CAAP: Pilot Year 1 1988-89, Collegiate Assessment of Academic Proficiency, American College Testing Program, 1989.

ACT, Review of COMP Development: Instrument Design and Evaluation of the Objective Test for Use in Program Evaluation. College Outcome Measures Program (COMP), American College Testing Program, 1990.

ACT, COMP 1987-88. College Outcome Measures Program (brochure), American College Testing Program, 1987.

Association of American Colleges. Integrity in the College Curriculum: A Report to the Academic Community. 1985, Washington, D.C.

Forrest, Aubrey and Steele, Joe M. Defining and Measuring General Educaton Knowledge and Skills. College Outcome Measures Project, American College Testing Program, 1982.

Brookfield, Stephen D. Developing Critical Thinkers, 1987, San Francisco, CA Jossey-Bass Inc. Publishers.

Ennis, R. H. A Logical Basis for Measuring Critical Thinking Skills. Educational Leadership, 1985, 43(2): 44-48.

Glaser, E. M. Critical Thinking: Education for Responsible citizenship in a Democracy. National Forum, 1985, 65(1): 24-27.

McMillan, James H. Enhancing College Students Critical Thinking: A Review of Studies. Research in Higher Education, 1987, Vol. 26, 3-29.

Moore, Brooke Noel and Parker, Richard. Critical Thinking: Evaluating Claims and Arguments in Everyday Life. 2nd ed. 1989, Mountain View, CA: Mayfield Publishing Company.

National Institute of Education. Involvement in Learning: Realizing the Potential of American Higher Education. 1984, Washington, D.C.: U.S. Government Printing Office.

Postman, Neil. Critical Thinking in the Electronic Age. National Forum, 1985, LXV(1), 4-8.

Sternberg, R. J. Teaching Critical Thinking, Part 2: Possible Solutions. Phi Delta Kappan, 1985, 67(4): 277-280.

Weinstein, Mark. Reason and Critical Thinking. Informal Logic, 1988, X(1), 1-20.